J Biostat Epidemiol. 2017; 3(2): 41-8.

Original Article

Comparing zero-inflated Poisson, Poisson gamma, and Poisson lognormal regression models in dental health data

Farzaneh Neysari[1], Abbas Bahrampour[1*], Yunes Jahani[1]

[1] Modeling in Health Research Center, Institute for Futures Studies in Health AND Department of Biostatistics and Epidemiology, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran

ARTICLE INFO

Received 19.02.2017
Revised 19.02.2017
Accepted 27.04.2017
Published 04.06.2017

Key words:
Poisson distribution;
Regression;
Decayed, missing and filled teeth;
Decay-missing-filled index

ABSTRACT

Background & Aim: Statistical modeling is one of the most suitable methods for analyzing the relationship between health and medical issues. In the situation of analysis of zero-inflated data, there are different methods. In this study, the models Poisson, Poisson gamma, and Poisson lognormal regression were compared.

Methods & Materials: This cross-sectional study was conducted to determine the influential factors on decay-missing-filled (DMF) index by the three mentioned models using the data of 808 first-grade children of the primary school in Kerman, Iran. The command PROC NLMIXED in SAS software was applied for fitting the models on data. For comparing the models, we applied the Akaike's criterion (AIC), mean square error (MSE) criterion and confidence interval (CI).

Results: The AIC and CI showed that the Poisson lognormal model was better than the others due to a level of significance. The variables of the students' place of living, mothers' jobs, fathers' jobs, the region, sex, optic problems, and behavioral problems had a significant effect on DMF index.

Conclusion: Poisson lognormal was better than the other models in dental health data.

Introduction

Oral diseases and more specifically dental caries are the most prevalent health problems. In this regard, oral health can contribute to improving such health conditions greatly which has been added to Iran's healthcare system since 1993 (1).

Zero-inflated Poisson (ZIP) regression and zero-inflated Poisson gamma (ZIPG) models were surveyed in a study by Lord et al. (2). Another study by Nodtvedt et al. studied ZIPG model (3). In a study carried out by Renner et al., ZIP regression and zero-inflated Poisson lognormal (ZIPLN) models were pinpointed (4). Bonate et al. (5) and Fang (6) did studies on ZIP regression and ZIPG models, respectively.

Zero-inflated models were first introduced by Lambert (7). Such models are used when the data contain too many zeros, especially when models such as Poisson and Poisson gamma are not effective for modeling. The most commonly used models are ZIP and ZIPG, respectively (7).

The zero-inflated models are mixed models which combine a component count with a binary component. The zero will occur in both single parts. The first part with the probability of p and the second part with the probability of 1– p, which the second one includes the Poisson distribution. The zeros for the first part are structural zeros and for the second one from the Poisson distribution are called sampling zeros

* Corresponding Author: Abbas Bahrampour, Postal Address: Department of Biostatistics and Epidemiology, Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
Email: yonesjahani@yahoo.com

Please cite this article in press as: Neysari F, Bahrampour A, Jahani Y. Comparing zero-inflated Poisson, Poisson gamma and Poisson lognormal regression models in dental health data. *J Biostat Epidemiol. 2017; 3(2): 41-8*

(8). The probability mass function for the zero-inflated are (9):

$$f(Y=y)=\begin{cases} p+(1-p)\,g(0) & \text{for } y=0 \\ (1-p)g(y) & \text{for } y=1,2,3,\ldots \end{cases}$$

In which g (0) is the count model probability function and $p\in[0,1]$ is the probability of having too many zeros.

ZIP regression model is one of the most commonly used models among zero-inflated models. In this model, the response variable has too many zeros which are composed of two processes. The first one is administered using a binary distribution that generates zero. The second one is administered using a Poisson distribution which generates the numbers which are likely to have too many zeros (10). The zero-inflated model is defined as follows (11):

$$f(Y=y)=\begin{cases} \Pr(Y_i=0)=p+(1-p)\,e^{-\lambda} \\ \Pr(Y_i>0)=(1-p)\dfrac{\lambda^{y_i}e^{-\lambda}}{y_i!} \end{cases}$$

In which Y is the response variable and X the independent variable. In this model, the mean is $(1-p)\lambda$ and the variance is $\lambda\,(1-p)\,(1+\lambda p)$ (10), we also have:

$$\ln\lambda=\beta_0+\sum\beta_i x_i$$
$$\text{Logit (p)} = \log\left(\frac{p}{1-p}\right)=\gamma_0+\sum\gamma_i x_i$$

Log-likelihood function in ZIP model for each individual is defined as follows (12):

$$\text{if } (Y=0) \text{ then } \log(\text{likelihood}) = \log(p+(1-p)\exp(-\lambda))$$
$$\text{if } (Y>0) \text{ then } \log(\text{likelihood}) = \log(1-p)+y\log(\lambda)-\lambda-\log(y!)$$

ZIPG regression model is defined using a mixed distribution, in which p is the probability of having too many zeros and 1 - p is allocated to the Poisson gamma distribution, in which $0\le p\le 1$.

The Poisson gamma distribution is a mixture of a mixed Poisson distribution in which the Poisson distribution mean λ has a gamma distribution which causes overdispersion in this model (13).

Consequently, the density function of ZIPG regression model will be as follows (13):

$$f(Y=y)\begin{cases} p+(1-p)(1+\alpha\lambda)^{-\frac{1}{\alpha}} & ,\ y=0 \\ (1-p)\dfrac{\Gamma\left(y+\frac{1}{\alpha}\right)}{y!\,\Gamma\left(\frac{1}{\alpha}\right)}(1+\alpha\lambda)^{-\frac{1}{\alpha}}\left(\dfrac{\alpha\lambda}{1+\alpha\lambda}\right)^{y} & ,\ y=1,2,3,\ldots \end{cases}$$

In which α is the dispersion parameter. The mean and the variance of the distribution of Poisson gamma inflated in zero will be as follows (13):

$$E(Y)=(1-p)\,\lambda \text{ و } \operatorname{var}(Y)=(1-p)\,\lambda\,(1+p\lambda+\alpha\lambda)$$

ZIPG regression model is the same as the ZIP regression model with random effects. Its random effect which is shown with exp (u) has a gamma distribution with ν *and* ν parameters (14). So we have:

$$y|\lambda \sim \text{poisson}(\lambda)$$
$$\ln\lambda=\beta_0+\sum\beta_i x_i+u$$
$$\exp(u)\sim \text{gamma}(\nu,\nu)$$
$$\text{Logit (p)} = \log\left(\frac{p}{1-p}\right)=\gamma_0+\sum\gamma_i x_i$$

Log-likelihood function in ZIPG model for each individual is defined as follows:

$$\text{if } (Y=0) \text{ then } \log(\text{likelihood}) = \log\left(p+(1-p)f(Y=y)\right)$$
$$\text{if } (Y>0) \text{ then } \log(\text{likelihood}) = \log((1-p)f(Y=y))$$

In the above equation, $f(Y=y)$ is the same as ZIPG distribution which is defined earlier.

ZIPLN model is also a mixed model which its parameters are more complex than Poisson regression model and Poisson gamma. The reason is that the Poisson Lognormal model does not have a closed form. This model can be defined as follows (15):

$$y|\lambda \sim \text{poisson}(\lambda)$$
$$\ln\lambda=\beta_0+\sum\beta_i x_i+u$$
$$\exp(u)\sim lognormal(0,\sigma^2)$$
$$\text{Logit (p)} = \log\left(\frac{p}{1-p}\right)=\gamma_0+\sum\gamma_i x_i$$

Poisson lognormal probability density function is as follows (16):

$$P_y=\int_0^{\infty}P(y|\lambda)f(\lambda)d\lambda$$

And we finally have (17):

$$P_y = \int_0^\infty \left(\frac{\lambda^y e^{-\lambda}}{y!}\right)\left\{\frac{1}{\lambda\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln\lambda - \mu)^2}{2\sigma^2}\right)\right\} d\lambda$$

Poisson lognormal model mean and variance are as follows (17):

$$E(Y_i) = \mu_i \exp(0.5\sigma^2)$$
$$\text{Var}\,(Y_i) = E(Y_i) + [E(Y_i)]^2(\exp(\sigma^2) - 1)$$

As a result, the ZIPLN density function will be as follows:

$$f(Y = y) = \begin{cases} p + (1-p)p(0) \\ \quad (1-p)P_y \end{cases}$$

The ZIPG regression model had a closed form which finally ended in a negative binomial model while the ZIPLN regression model did not have a closed form and that is why its parameters estimating are more complex. On the other hand, the ZIPLN regression model is as same as the ZIP model with random effects. These random effects have a normal distribution (14). The log-likelihood function of this distribution for each individual is defined as follows:

$$\text{if}\,(Y = 0) \quad \text{then}\ \ln\,(\text{likelihood}) = \log(p + (1-p)f(Y = 0))$$
$$\text{if}\,(Y > 0) \quad \text{then}\ \ln\,(\text{likelihood}) = \log\ (1-p) + \log(f(Y = y))$$

The equality of mean and variance in Poisson regression is the most important feature of this model (18). But, in statistical modeling, the lack of equality between the variance and the mean in applying the Poisson model causes misinterpretation of the model's parameters (19). According to the observed data in this study, models such as Poisson gamma, Hurdle, and the zero-inflated models are introduced as an alternative to the Poisson model (20).

Zero-inflated models were surveyed according to the data of this study. ZIP, ZIPG, zero-inflated Poisson normal (ZIPN), zero-inflated generalized Poisson (ZIGP), ZIPLN and zero-inflated Binomial (ZIB) are among the introduced zero-inflated regression models. In this study, the ZIP, ZIPG, and ZIPLN were compared. The ZIPLN model discussed in this study is scarcely surveyed in most studies on zero-inflated models.

Study of DMF indices in 6-7-year-old children, which is one of the target groups in the healthcare system, is considered in this study which is a count type variable.

Methods

This cross-sectional study was performed on 808 primary school children in Kerman, Iran. Kerman is divided into five districts namely North, South, East, West, and Center. Three schools were randomly selected from each district. Then, the first grade of primary school children was selected and subsequently, their information from their health file, which had been provided by their health teacher, was obtained. The categorized variables were entered as indicator variables. After that, the ZIP, ZIPG and ZIPLN were fitted using PROC NLMIXED in SAS software 9.2 Version.

The significance level in this study was 0.050. In the zero-inflated part, logit link function was taken into account. In the three final models, first, all of the variables were entered into the model and the variables whose P was less than or equal to 0.2 were selected for the multiple analysis. Then, applying backward elimination of the regression models, the ZIP, ZIPG and ZIPLN were fitted. Finally, these three models were compared. As in the count models the connection function is logarithmic, the incidence rate ratio (IRR) was used for the interpretation of the coefficients.

In this study, the Akaike's criterion (AIC) was used to compare the efficacy of models. AIC explains an equation between fit and complexity. AIC is defined generally as below:

AIC = 2K-2lnL, in which, K is the number of parameters and L is the maximum likelihood of the models. Among the models, the one was prioritized which had the least AIC. If AIC had very little difference between the two models and they were very similar, the confidence intervals (CI) were compared and the model with narrower CI was selected.

Results

24% of the participants were girls and 76% were boys. 97% of the students were living with their parents. Almost 39.1% of the fathers and

45% of the mothers held diploma. 73.1% of the mothers were housekeepers and 48.1% of the fathers were self-employed. Almost all the participants had healthy vision and common sense with no problem in this respect. As in the count models the connection function is logarithmic, IRR was used for the interpretation of the coefficients.

In table 1, the frequency of the independent variables is reported.

Table 1. The distribution of the Independent variables in the first grade of primary school students in Kerman, Iran

Variable	Frequency (%)
Place of living	
With parents	784 (97.0)
With Mother	19 (2.4)
With Father	0 (0)
With Other	5 (0.6)
Mother's Education	
Illiterate and Primary school	92 (11.4)
Guidance school	141 (17.5)
High school and Diploma	364 (45.0)
University	211 (26.1)
Father's Education	
Illiterate and Primary school	94 (11.6)
Guidance school	155 (19.2)
High school and Diploma	316 (39.1)
University	243 (30.1)
Mother's Job	
Employee	172 (21.3)
Housekeeper	591 (73.1)
Teacher	45 (5.6)
Father's Job	
Employee	296 (36.6)
Free	389 (48.1)
Working	53 (6.6)
Teacher	46 (5.7)
Military	24 (3.0)
School Type	
Private school	173 (21.4)
Governmental	635 (78.6)
Parent relationship	
Yes	249 (30.8)
No	559 (69.2)
The region	
Area 1	454 (56.2)
Area 2	354 (43.8)
Sex	
Girl	194 (24)
Boy	614 (76)
Optic problem	
Yes	68 (8.4)
No	740 (91.6)
Behavior Problem	
Yes	23 (2.8)
No	785 (97.2)

In table 2, the distribution of the decay-missing-filled (DMF) response variable has been reported, which 46.8% of the whole data was formed by zero.

Table 2. The distribution of the decay-missing-filled (DMF) response variable in the primary school students of Kerman, Iran

DMF index	Frequency (%)
0	378 (46.8)
1	57 (7.1)
2	87 (10.8)
3	78 (9.7)
4	69 (8.5)
5	50 (6.2)
6	42 (5.2)
7	24 (3.0)
8	23 (2.8)
Total	808 (100)

DMF: Decay-missing-filled

In tables 3-5, the obtained ZIP, ZIPG and ZIPLN regression models have been reported along with the models' coefficients, the standard error, the odds ratio (OR) index and its CI and P for the logistic part of the model. For the Poisson part of the model the model's coefficients, the standard error, the IRR index along with its CI and P have been reported.

In table 6, the AIC for the ZIP r, ZIPG and ZIPLN regression models have been reported. The AIC of the ZIP model has been 2.5 units greater than the one of the other two models which were eliminated from the best models. For ZIPG and ZIPLN models, a similar AIC has been obtained. Therefore, the width of CI and mean square error (MSE) criterion were compared in table 7. A model with shorter CI was preferred. Hence, both CI and MSE criterion for ZIPLN was less than the other models.

Discussion

The Poisson regression model is one of the most widely-used models for count data such as the number of opium withdrawal, the number of decayed teeth, etc (12). The Poisson regressions are a part of statistical models which are called generalized linear models (GLM) (12). Sometimes when these count data have too many zeros, using a Poisson regression model will no longer be adequate (13).

Table 3. The relationship between the demographic variables and the decay-missing-filled (DMF) index using the zero-inflated Poisson (ZIP) model in the first-grade students of the primary schools of Kerman, Iran*

Variable		Logistic			Poisson		
		Parameter ± Standard Error	**OR (95% CI)**	**P-value**	**Parameter ± Standard Error**	**IRR (95% CI)**	**P-value**
Intercept		-4.84	-	< 0.001	0.91	-	< 0.001
Mother's Job	Housekeeper	-1.82 ± 0.39	0.16 (0.07-0.34)	< 0.001	0.31 ± 0.04	1.36 (1.24-1.49)	< 0.001
	Teacher	-1.37± 0.43	0.25 (0.10-0.59)	0.001	0.22 ± 0.09	1.25 (1.03-1.51)	0.010
	Employee	Reference	-	-	Reference	-	-
Place of living	With parent	-1.63 ± 0.39	0.19 (0.08-0.42)	< 0.001	0.37 ± 0.05	1.46 (1.31-1.62)	< 0.001
	With Mother	-1.07 ± 0.59	0.34 (0.10-1.09)	0.070	-0.03 ± 0.15	0.96 (0.71-1.29)	0.810
	With Other	Reference	-	-	Reference	-	-
Father's Job	Free	-	-	-	0.05 ± 0.05	1.06 (0.94-1.19)	0.310
	Working	-	-	-	0.23 ± 0.10	1.25 (1.02-1.54)	0.020
	Teacher	-	-	-	0.05 ± 0.13	1.05 (0.80-1.37)	0.710
	Military	-	-	-	0.01 ± 0.16	1.01 (0.74-1.39)	0.920
	Employee	-	-	-	Reference	-	-
Behavior problem	No	1.53 ± 0.63	4.63 (1.32-16.27)	0.010	-	-	-
	Yes	Reference	-	-	-	-	-
The region	Area 1	0.35 ± 0.15	1.43 (1.04-1.95)	0.020	-	-	-
	Area 2	Reference	-	-	-	-	-
Sex	Girl	1.28 ± 0.18	3.62 (2.51-5.21)	< 0.001	-	-	-
	Boy	Reference	-	-	-	-	-
Optic Problem	No	0.67 ± 0.29	1.96 (1.09-3.50)	0.020	-	-	-
	Yes	Reference	-	-	-	-	-

* Zero-inflated Poisson model

-2Log-likelihood = 2792.5; Akaike's criterion (AIC) = 2828.5; OR: Odds ratio; CI: Confidence interval; IRR: Incidence rate ratio

In such cases, application of a Poisson model will result in the wrong estimation of the model's parameter (13). In such cases, the zero-inflated models should be applied. Overdispersion results in the underestimation of the standard error and this increases the test statistics which leads to a false significance of a variable (19).

Table 4. The relationship between the demographic variables and the decay-missing-filled (DMF) index using the zero-inflated Poisson gamma (ZIPG) model in the first-grade students of the primary schools of Kerman, Iran*

Variable		Logistic			Poisson		
		Parameter ± Standard Error	**OR (95% CI)**	**P-value**	**Parameter ± Standard Error**	**IRR (95% CI)**	**P-value**
Intercept		-5.47	-	< 0.001	0.89	-	< 0.001
Mother's Job	Housekeeper	-2.02 ± 0.44	0.13 (0.05-0.31)	< 0.001	0.30 ± 0.05	1.35 (1.23-1.49)	< 0.001
	Teacher	-1.66 ± 0.49	0.18 (0.07-0.49)	0.001	0.21 ± 0.10	1.24 (1.01-1.53)	0.040
	Employee	Reference	-	-	Reference	-	-
Place of living	With parent	-1.77 ± 0.44	0.16 (0.07-0.40)	< 0.001	-0.03 ± 0.16	0.96 (0.69-1.34)	0.830
	With Mother	-1.14 ± 0.62	0.31 (0.09-1.07)	0.060	00.37 ± 0.06	1.45 (1.29-1.63)	< 0.001
	With Other	Reference	-	-	Reference	-	-
Father's Job	Free	0.29 ± 0.18	1.34 (0.94-1.91)	0.100	0.07 ± 0.06	1.07 (0.94-1.22)	0.280
	Working	0.43 ± 0.33	1.54 (0.79-2.99)	0.190	0.24 ± 0.11	1.28 (1.01-1.61)	0.030
	Teacher	0.72 ± 0.36	2.06 (1.01-4.21)	0.040	0.07 ± 0.15	1.08 (0.80-1.45)	0.600
	Military	-0.01 ± 0.47	0.98 (0.38-2.48)	0.970	0.02 ± 0.18	1.02 (0.71-1.45)	0.900
	Employee	Reference	-	-	Reference	-	-
Behavior problem	No	1.73 ± 0.73	5.68 (1.33-24.28)	0.010	-	-	-
	Yes	Reference	-	-	-	-	-
The region	Area 1	0.39 ± 0.16	1.48 (1.07-2.04)	0.010	-	-	-
	Area 2	Reference	-	-	-	-	-
Sex	Girl	1.29 ± 0.18	3.65 (2.52-5.30)	< 0.001	-	-	-
	Boy	Reference	-	-	-	-	-
Optic Problem	No	0.72 ± 0.30	2.07 (1.12-3.80)	0.010	-	-	-
	Yes	Reference	-	-	-	-	-

* Zero-inflated Poisson gamma model

Overdispersion α (%95 CI) = 0.06 (0.006,0.11), P-value < 0.001; -2Log-likelihood = 2780; Akaike's criterion (AIC) = 2826; OR: Odds ratio; CI: Confidence interval; IRR: Incidence rate ratio

Table 5. The relationship between the demographic variables and the decay-missing-filled (DMF) index using zero-inflated Poisson lognormal (ZIPLN) model in the first-grade students of the primary schools of Kerman, Iran*

Variable		Logistic			Poisson		
		Parameter ± Standard Error	OR (95% CI)	P-value	Parameter ± Standard Error	IRR (95% CI)	P-value
Intercept		-5.47	-	< 0.001	0.87	-	< 0.001
Mother's Job	Housekeeper	-2.01 ± 0.43	0.13 (0.05-0.31)	< 0.001	0.30 ± 0.05	1.35 (1.23-1.49)	< 0.001
	Teacher	-1.65 ± 0.48	0.19 (0.07-0.49)	0.001	0.21 ± 0.10	1.23 (1.01-1.52)	0.040
	Employee	Reference	-	-	-	-	-
Place of living	With parent	-1.76 ± 0.43	0.31 (0.09-1.07)	0.060	-0.03 ± 0.05	1.44 (1.28-1.62)	0.820
	With Mother	-1.14 ± 0.61	0.17 (0.07-0.40)	< 0.001	-0.03 ± 0.16	0.96 (0.69-1.33)	< 0.001
	With Other	Reference	-	-	-	-	-
Father's Job	Free	0.29 ± 0.17	1.33 (0.94-1.90)	0.100	0.07 ± 0.06	1.07 (0.94-1.22)	0.280
	Working	0.43 ± 0.33	1.53 (0.79-2.97)	0.190	0.24 ± 0.11	1.28 (1.01-1.60)	0.030
	Teacher	0.72 ± 0.36	2.05 (1.01-4.19)	0.040	0.05 ± 0.07	1.08 (0.80-1.44)	0.600
	Military	-0.02 ± 0.47	0.97 (0.38-2.45)	0.970	0.01 ± 0.17	1.02 (0.71-1.44)	0.910
	Employee	Reference	-	-	-	-	-
Behavior problem	No	1.72 ± 0.72	5.62 (1.34-23.49)	0.010	-	-	-
	Yes	Reference	-	-	-	-	-
The region	Area 1	0.39 ± 0.16	1.48 (1.07-2.03)	0.010	-	-	-
	Area 2	Reference	-	-	-	-	-
Sex	Girl	1.29 ± 0.18	3.65 (2.51-5.29)	< 0.001	-	-	-
	Boy	Reference	-	-	-	-	-
Optic Problem	No	0.72 ± 0.30	0.004 (0-0.04)	< 0.001	-	-	-
	Yes	Reference	-	-	-	-	-

* Zero inflated Poisson lognormal model

Variance (u) (%95 CI) = 0.054 (0.004,0.10), P-value < 0.0001; -2Log Likelihood = 2780.7; Akaike's criterion (AIC) = 2826.7; OR: Odds ratio; CI: Confidence interval; IRR: Incidence rate ratio

Table 6. The statistical indices of the models under question in brief

Models	AIC	-2log-likelihood
ZIP	2828.5	2792.5
ZIPG	2826	2780
ZIPLN	2826.7	2780.7

AIC: Akaike's criterion; ZIP: Zero-inflated Poisson; ZIPG: Zero-inflated Poisson gamma; ZIPLN: Zero-inflated Poisson lognormal

One way to deal with overdispersion is to use mixed Poisson gamma models. Another way is to use the zero-inflated models which were used in this study because of having too many zeros in the data. In order to control the dispersion, ZIP, ZIPG and ZIPLN regression models were applied among which the ZIPLN model was preferred to the other ones.

Table 7. The width of confidence intervals of the odds ratio (OR) and IRR indices and mean square error (MSE) criterion for the significant parameters in zero-inflated Poisson gamma (ZIPG) and zero-inflated Poisson lognormal (ZIPLN) regression models

Variables		ZIPG		ZIPLN	
		Logistic (OR)	Poisson gamma (IRR)	Logistic (OR)	Poisson lognormal (IRR)
Intercept		0.04	0.46	0.04	0.46
Mother's Job	Housekeeper	0.25	0.26	0.25	0.26
	Teacher	0.42	0.52	0.42	0.52
	Employee	-	-	-	-
Place of living	With Parent	0.33	0.34	0.33	0.33
	With Mother	0.98	0.64	0.97	0.64
	With Other	-	-	-	-
Father's Job	Free	0.97	0.28	0.96	0.27
	Working	2.19	0.59	2.18	0.58
	Teacher	3.20	0.64	3.18	0.64
	Military	2.09	0.74	2.06	0.73
	Employee	-	-	-	-
The region		0.96	-	0.96	-
Sex		2.78	-	2.77	-
Optic problem		2.67	-	2.65	-
Behavior problem		22.95	-	22.15	-
MSE		6.18	-	6.08	-

ZIPG: Zero-inflated Poisson gamma; ZIPLN: Zero-inflated Poisson lognormal; OR: Odds ratio; IRR: Incidence rate ratio; MSE: Mean square error

Cheng et al. stated how to use ZIP models for fitting the data with too many zeros, which was taken into account in this study (21). Aguero-Valverde showed that Poisson gamma regression models, Poisson lognormal regression models, ZIPG and ZIPLN have been fitted for the data. Mean standard error of the models is one of the factors used for the comparison of the models (14). In this study, the ZIPLN models are preferred to the others. In another study done by Renner et al. Poisson, Poisson lognormal, ZIP, and ZIPLN regression models were fitted to the data. In this study, the ZIPLN model was preferred to other ones including the ZIP model (4).

Conclusion

Zero-inflated models are used for the count data with too many zeros. ZIP and ZIPG models are the most commonly used models. In this study, ZIPLN model was reported. The lognormal Poisson model in similar studies has been preferred to Poisson model and Poisson gamma regression models. In this study, these three models were also reported based on their AIC as well as the width of the CI of the model's parameters and MSE criterion. Finally, The ZIPLN model was preferred to the others.

Conflict of Interests

Authors have no conflict of interests.

Acknowledgments

Hereby we do appreciate all people and our colleagues who helped us with their science efforts to develop this paper.

References

1. Saied-Moallemi Z, Virtanen JI, Ghofranipour F, Murtomaa H. Influence of mothers' oral health knowledge and attitudes on their children's dental health. Eur Arch Paediatr Dent 2008; 9(2): 79-83.
2. Lord D, Washington SP, Ivan JN. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. Accid Anal Prev 2005; 37(1): 35-46.
3. Nodtvedt A, Sanchez J, Dohoo I. Applying a zero-inflated negative binomial model to hierarchical count data. Proceedings of the 10th Symposium of the International Society for Veterinary Epidemiology and Economics; 2003 Nov. 17-21; Vina del Mar, Chile; 2003. p. 307 [Online]. [cited Nov. 2003]; Available from: URL: http://www.sciquest.org.nz/node/63163.
4. Renner HM, Reynolds JH, Sims M, Renner M. Evaluating the power of surface attendance counts to detect long-term trends in populations of crevice-nesting auklets. Environ Monit Assess 2011; 177(1-4): 665-79.
5. Bonate PL, Sung C, Welch K, Richards S. Conditional modeling of antibody titers using a zero-inflated poisson random effects model: Aapplication to Fabrazyme. J Pharmacokinet Pharmacodyn 2009; 36(5): 443-59.
6. Fang R. Zero-inflated negative binomial (ZINB) regression model for over-dispersed count data with excess zeros and repeated measures, an application to human micro biota sequence data [MSc Thesis]. Tangshan, China: North China Coal Medical University; 2008.
7. Lambert D. Zero-inflated poisson regression, with an application to defects in manufacturing. Technometrics 1992; 34(1): 1-14.
8. Pahel BT, Preisser JS, Stearns SC, Rozier RG. Multiple imputation of dental caries data using a zero-inflated Poisson regression model. J Public Health Dent 2011; 71(1): 71-8.
9. Erdman D, Jackson L, Sinko A. Zero-inflated Poisson and zero-inflated negative binomial models using the countreg procedure. Cary, NC: SAS Institute Inc; 2009.
10. Hall DB. Zero-inflated Poisson and binomial regression with random effects: A case study. Biometrics 2000; 56(4): 1030-9.
11. Millar RB. Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes' factors. Biometrics 2009; 65(3): 962-9.
12. Flynn M. More flexible GLMs zero-inflated models and hybrid models [Online]. [cited 2009]; Available from: URL: https://www.casact.org/pubs/forum/09wforu

m/flynn-francis.pdf.
13. Chow NT, Steenhard D. A flexible count data regression model using SAS- PROC NLMIXED. Denver, CO: SAS Global Forum: Statistics and Data Analysis; 2009.
14. Aguero-Valverde J. Full Bayes Poisson gamma, Poisson lognormal, and zero-inflated random effects models: Comparing the precision of crash frequency estimates. Accid Anal Prev 2013; 50: 289-97.
15. Staub KE, Winkelmann R. Consistent estimation of zero-inflated count models. Health Econ 2013; 22(6): 673-86.
16. Izsak R. Maximum likelihood fitting of the Poisson lognormal distribution. Environ Ecol Stat 2008; 15(2): 143-56.
17. Williams MS, Ebel ED. Methods for fitting the Poisson-lognormal distribution to microbial testing data. Food Control 2012; 27(1): 73-80.
18. Ridout M, Dem´etrio CG, Hinde J. Models for count data with many zeros. Proceedings of the International Biometrics Conference; 1998 Dec. 13-18; Cape Town, South Africa.
19. Schmidt AM, Pereira BM, Vieira PP. Do we always need a zero-inflated model to capture an apparent excess of zeros? [Online]. [cited 2008]; Available from: URL: www.dme.im.ufrj.br/arquivos/publicacoes/arquivo211.pdf.
20. Ngatchou-Wandji J, Paris C. On the Zero-Inated Count Models with Application to Modeling Annual Trends in Incidences of Some Occupational Allergic Diseases in France. J Data Sci 2011; 9: 639-59.
21. Cheng J, Cheng NF, Guo Z, Gregorich S, Ismail AI, Gansky SA. Mediation analysis for count and zero-inflated count data. Stat Methods Med Res 2016; 962280216686131.